Plagiarism detection in software using efficient string matching

Authors: Kusum Lata Pandey — Ewing Christian College, Allahabad, India

Suneeta Agarwal — Motilal Nehru National Institute of Technology, Allahabad, India

Sanjay Misra — Atilim University, Ankara, Turkey

Rajesh Prasad — Ajay Kumar Garg Engineering College, Ghaziabad, India

2012 Article

Bibliometrics

Published in:

· Proceeding
ICCSA'12 Proceedings of the 12th international conference on Computational Science and Its Applications - Volume Part IV
Pages 147-156
Springer-Verlag Berlin, Heidelberg ©2012
table of contents ISBN: 978-3-642-31127-7 doi>10.1007/978-3-642-31128-4_11

String matching refers to the problem of finding occurrence(s) of a pattern string within another string or body of a text. It plays a vital role in *plagiarism detection*in software codes, where it is required to identify similar program in a large populations. String matching has been used as a tool in a software metrics, which is used to measure the quality of software development process. In the recent years, many algorithms exist for solving the string matching problem. Among them, Berry---Ravindran algorithm was found to be fairly efficient. Further refinement of this algorithm is made in TVSBS and SSABS algorithms. However, these algorithms do not give the best possible shift in the search phase. In this paper, we propose an algorithm which gives the best possible shift in the search phase and is faster than the previously known algorithms. This algorithm behaves like Berry-Ravindran in the worst case. Further extension of this algorithm has been made for parameterized string matching which is able to detect plagiarism in a software code.